

Positioning Evaluation and Ground Truth Definition for Real Life Use Cases

Carlos Martínez de la Osa, Grigorios G. Anagnostopoulos, Mauricio Togneri,
Michel Deriaz and Dimitri Konstantas

Information Science Institute
GSEM/CUI

University of Geneva
Geneva, Switzerland

{carlos.martinez, grigorios.anagnostopoulos, mauricio.togneri}@unige.ch
{michel.deriaz, dimitri.konstantas}@unige.ch

Abstract—Evaluating positioning systems has become a matter of heated debate during the last years. There is no clear standard on how these technologies should be evaluated, and no predominant solution for defining the ground truth in order to compare the position estimates. In this paper, we propose a simple and inexpensive solution for tackling both of these problems in real life use cases. With the proposed methodology, it is possible to measure both static and moving targets, by creating a predefined path with checkpoints. Then, a tester, walking over them, while moving or standing still, indicates when the device was over the aforementioned checkpoints. It is also specified how to evaluate the estimates by comparing them with interpolated points of the ground truth trajectory. Two methods are proposed for performing such interpolation. Finally, in order to evaluate the performance of the positioning system as well as the perceived utility of the position estimates from the end user’s point of view, a series of statistical parameters is discussed. Additionally, in the context of perceived utility by the end user, a parameter that measures the occurrence of abrupt changes in the position estimates is proposed.

Keywords—*Indoor Positioning; Outdoor Positioning; Ground truth definition; Positioning evaluation, Tracking.*

I. INTRODUCTION

In the last years, the indoor positioning field has experienced a very notable progress. The rise of technologies such as Bluetooth, Near Field Communication (NFC) or ultrasound, has allowed deployment costs to be reduced, increasing the research related to this field. At the same time, the way that these positioning systems are evaluated has gained more importance.

In this work, we describe a methodology for evaluating positioning systems in a predefined path and gathering the ground truth data which is used as reference in that evaluation.

The motivation of this work comes from the necessity of our research group to evaluate the performance of our positioning systems when utilized by moving end users.

Until now, the characteristics of our positioning systems have been measured using “static evaluation”. This method

consists in determining a point with known coordinates as ground truth, and compare it with the position estimates obtained with the positioning system at that location, without moving. It is commonly used in some indoor positioning competitions [1]. Also, an improved version of this concept was used at the last Indoor Positioning and Indoor Navigation (IPIN) conference competition [2]. In the work described by Pulkkinen and Verwijnen [3], the authors discuss about this method and propose some metrics to improve the scientific value of the evaluations, introducing interesting parameters like “environment-normalized error” or “shortest path error”.

The work presented by Schwartz, et al. [4] introduces three categories for the evaluation methods:

- 1) *Static evaluation*: described in the previous paragraph. This evaluation method can also be achieved with our proposed methodology, although it is not the main goal.
- 2) *Dynamic evaluation with predefined geometrical paths*: specific paths are defined in advance in a test field, and then followed by a person while using the positioning system under evaluation. We will refer to this person as the tester during the rest of this paper.
- 3) *Dynamic evaluation using a reference positioning system*: a positioning system with higher accuracy is used as a reference for the evaluation of the target system.

The third kind of evaluation methodology is widely used for outdoors positioning systems, using as a reference the Global Positioning System (GPS) which already present errors, typically, between 1 and 15 metres, depending on the quality of the system used and the place of measurement, or Differential GPS (DGPS) for more accurate measurements, around tens of centimetres. This method is also used indoors [4], [5], [6], although it is generally difficult or expensive to deploy. An exception can be found in the work proposed by Schmitt, et al. [7], where the authors use off-the-shelf components to build a robot able to gather the ground truth data necessary to the evaluation of a positioning system. This solution, nevertheless, presents some limitations, as the robot can not move in all surfaces, and it is not able to climb stairs.

The work described in this study lies under the second category, *dynamic evaluation with predefined geometrical paths*. A tester will define the ground truth from a predefined path,

This work was co-funded by the State Secretariat for Education, Research and Innovation of the Swiss federal government and the European Union, in the frame of the EU AAL project SmartHeat (aal-2014-153).

while, at the same time, recording the sequence of position estimates given by the positioning system. This type of measuring offers the possibility of introducing some metrics about the trajectory, or the overall shape of the route of a moving target, as mentioned in [3]. Improving these values could result in a more natural tracking behaviour in the positioning system for the end users.

Additionally, in this paper we also discuss the way of evaluating the positioning systems, once the data has been gathered. Many studies have focused on this topic, with interesting benchmarks proposed like the EVARILLOS benchmarking platform [8]. The parameters mentioned in these studies can be measured with our methodology. In another interesting work about Benchmarking Radio Frequency (RF)-based indoor localization [9], the authors focus on the reproducibility by controlling the RF interference and using a robot as a tester. Some other efforts in this direction can be found in the work done by Adler, et al. [10], where the authors offer an open virtual testbed for indoor localization.

Finally, in this work it is also highlighted the intention to measure the perceived utility from the end user's point of view. For this purpose, a new metric comparing the predefined and the estimated paths is proposed. Furthermore, we pay special attention to a parameter also considered crucial for the end user's perception of the system, which is the claimed accuracy of the position estimates. While outdoor position providers, like the GPS, provide a claimed accuracy of the position estimates, the concept is not so often discussed in indoor positioning and the related research. It is important to evaluate positioning providers not only by the quality of the positions estimates that they give, but also by how representative is the accuracy that they claim for each position estimate. This parameter is extremely important in heterogeneous positioning systems which try to use the best technology available [11], [12].

In a very useful survey done by Adler, et al. [13], the authors state that a high percentage of the authors, from the last five IPIN conferences, describe their methods of ground truth data gathering poorly or they do not describe them at all. One of the main goals of the current work is to offer a clear view of the evaluation methods utilized in our lab, so that it can be used as a reference in future works of our research group, in order to back up the presented results. At the same time, we hope to contribute to the state of the art in the positioning evaluation methodologies.

The rest of this paper is organised as follows. In Section II, we present the ground truth definition methodology, explaining the path creation and data recording. The position estimate evaluation is shown in Section III. In Section IV, we illustrate the process described on the paper with a detailed example. Finally, future work directions along with conclusions drawn are presented in Section V.

II. GROUND TRUTH DEFINITION

The first part of the proposed methodology consists on defining how the ground truth data are gathered. These are later used as a reference to evaluate the position estimates.

As discussed in the introduction, our method lies inside the category of *dynamic evaluation with a predefined geometrical path*. One of the important characteristics of the proposed method is the addition of predefined checkpoints along the path, in order to be more precise with the time in which the user walked over each stretch of the route. A high density of checkpoints makes the methodology be more accurate in time with the position of the user.

Once the procedure of defining the path is finished, the user will travel the path, recording the position estimates received, and indicating the moment he steps over every checkpoint. After gathering the estimates and the ground truth data, an evaluation of the positioning system will be possible.

A. Path creation

The first step is to create a path which simulates the route in a typical usage scenario. The created paths should also try to explore most of the coverage area of the positioning system. This path consists of a list of positions, as real world coordinates (latitude and longitude), that will be followed one after the other, and will serve as checkpoints for the testers. In order to avoid errors at this point, it is indispensable that the maps used to create the path and to estimate the position of the users have the same geographic calibration.

The minimum requirement fields for a checkpoint are latitude and longitude. Apart from these, the checkpoint can be enriched by adding other parameters that describe with more detail a position, such as altitude, floor, room, etc.

B. Data recording

The second step required, prior to the evaluation, is to record the positioning system data along the predefined path.

The application needed for this purpose is very simple. It has, as an input, the list with the checkpoints created in the previous step. Besides, it has access to the position providers that we want to evaluate, defining a position provider as the logic that transform raw sensor data into position estimates. Finally, it needs a way to enable the tester to indicate the moment he is passing over a checkpoint.

Before starting the recording, the tester should clearly identify the predefined path with clear landmarks in the real scenario. If these landmarks are not precise enough to be identified, the tester should measure the distances in the reference maps and in the real life scenario and place signs for every checkpoint, such as clear numbered labels on the ground.

In order to start recording the data, the tester must be placed over the first checkpoint. Then, he should indicate it to the application and start walking towards the second checkpoint. At the same time, the system logs the time the user began the

path and starts recording the position estimates coming from the different positioning providers that the tester selected to record, saving the timestamp of each estimate at each time. Consecutively, the application will continue logging these estimates while the user is moving, the data being recorded at all times, not only at the checkpoints. The tester should indicate again every time he steps over a checkpoint, in order to save the time when he walked over each of them, until he reaches the end of the path where the application will stop gathering information.

In between two consecutive checkpoints the tester should walk at a steady pace over the line that links both checkpoints in order to maximize the accuracy of the method. This pace can be different during different stretches of the path as long as it remains constant in every stretch.

C. Procedure example

An example of a result can be seen in Figure 1, where we have created a path going from the inside of the building Battelle A in the University of Geneva to a point outside, in the park.



Figure 1. Example of a predefined path

We desire to measure positioning algorithms programmed for Android mobile phones, so we created an Android application with two simple screens, as seen in Figure 2. In the first one, on the left, the tester can select the position providers he wants to record. Following, he will select the file containing the predefined path he wants to follow. Finally, the main interface is a screen with a button indicating the number of the next checkpoint, as shown in the right side of Figure 2.

III. POSITION ESTIMATE EVALUATION

Once the ground truth data has been gathered along with the position estimates, the evaluation of the results can be done. In this section, the different attributes of a position estimate will be described, as well as two methods to evaluate them, along with different parameters that can be evaluated with our procedure.

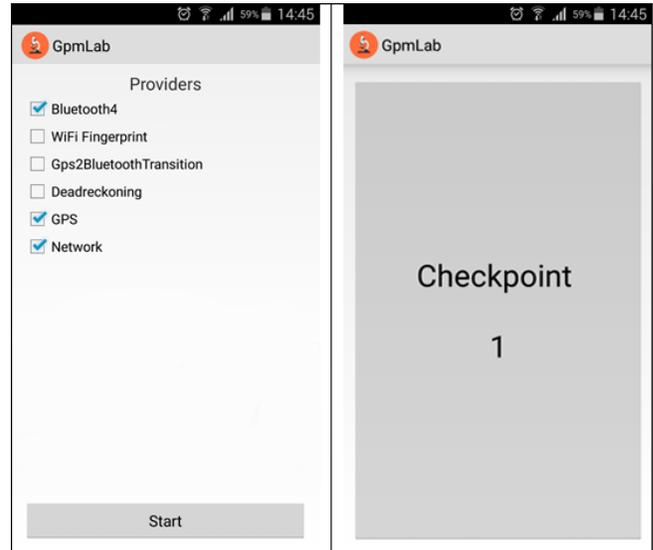


Figure 2. Two screenshots of our Android recording application

A. Position estimate

To begin with, it is necessary to establish the attributes that a position estimate must have in order to be suitable for evaluation. Therefore they are expected to be recorded by the position providers. These are:

- Latitude
- Longitude
- Provider name
- Timestamp

The two basic parameters that a position must have are the latitude and longitude coordinates, as they allow the identification of a specific point in the geographic coordinate system. Moreover, the name of the provider that estimated the position must be delivered, in order to give the user and the system information about the technology used. Finally, the timestamp is crucial as it allows to compare the ground truth and the estimated positions at a given time.

On top of this, a highly recommended parameter to calculate in a position provider is the claimed accuracy of the position estimate. As the position is not exact, but an estimate, it is needed to have an idea of the claimed quality of that estimate.

Additionally, the position estimate might also contain information about the altitude, the bearing of the user, etc. It can provide very useful information to record the room and the floor estimated by the position provider in order to measure the accuracy with which these are calculated. A more detailed description of the convenient properties that a position estimate can possess can be found in the work presented by Bekkelien et al. [14].

B. Ground truth linear interpolation

In order to evaluate a position provider, the position estimates will be compared with the ground truth data acquired. For this purpose, both of these data need to be known at

any particular point in time. As it is highly improbable that checkpoints of the ground truth and position estimates coincide exactly in time, interpolations from one or the other have to be created. To this end, two solutions are proposed. In both cases, the assumption made is that, when defining the ground truth using the application, the user was moving in a straight line at a steady pace between every two consecutive checkpoints.

1) *Estimate updates*: The first solution proposed will calculate an interpolated point in the ground truth path whenever a new position estimate is received. In order to do this, every time the algorithm processes a position estimate update, it checks its timestamp, and calculates the corresponding interpolated point in the ground truth path for this timestamp. This pair of points will later be compared in the evaluation phase. An example is shown in Figure 3, where the blue dots show the checkpoints and the time when the tester crossed them. The black triangles are the position estimates given by the position provider. Using the timestamps of those estimates, the interpolated checkpoints are created, shown as green squares, assuming a straight line and a steady pace between the two checkpoints. In this example, at the evaluation phase, the position estimates will be compared with the interpolated checkpoints created for that specific time.

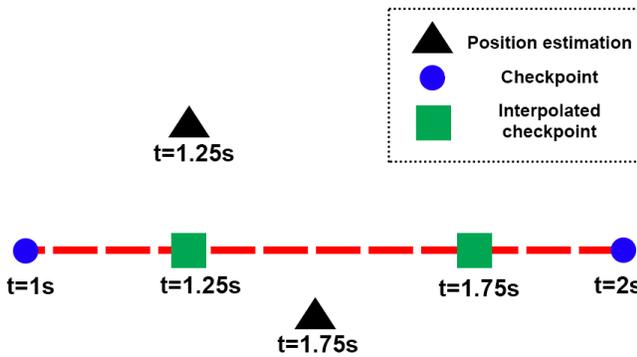


Figure 3. Linear interpolation example by estimate updates in a two-dimensional map

This approach is useful for evaluating each position estimate at the moment they are created, and checking the logic of the algorithm that calculates the position is accurate. On the other hand, there exists a drawback with this technique, in the case that a position provider is updating the position estimates with a very low frequency, only these points will be taken into account for the statistical analysis, but they will not be representative for the whole path. For example, a user starts recording a path which is one kilometre long, and the provider only gives two position updates, one at the beginning and another one at the end, both of them very accurate. Using this method, the evaluation will later conclude that the provider is very accurate, while most of the time that the user was utilizing the position provider, it was not giving new position updates.

2) *Fixed time interval*: In order to better evaluate the end user's perception of the positioning provider, a different

interpolation method is proposed. In this case, the tester will choose a fixed time interval (T). Then, the system will create an interpolated point in the ground truth periodically every T ($0, T, 2T, 3T, \dots$), including the beginning and the end of the path. These points will be compared to the most recent position estimate received at the corresponding timestamp. The shortest the time interval is, the more precise the evaluation will be, but it will proportionally increment the processing time. This method copes with the limitation of the first algorithm, as it compares the position of the tester very frequently with the estimates. Figure 4 shows an example of this procedure. As before, the blue dots show the checkpoints and the time when they were created. The black triangles are the position estimates given by the position provider. The green squares are the interpolated checkpoints created every $T=0.25s$, assuming a straight line and a steady pace between checkpoints. In this case, the interpolated checkpoints at $t=1.25s$ and $t=1.5s$ will be compared with the position estimate at $t=1.25s$, while the interpolated checkpoint at $t=1.75s$ and the checkpoint at $t=2s$ will be compared with the estimate at $t=1.75s$.

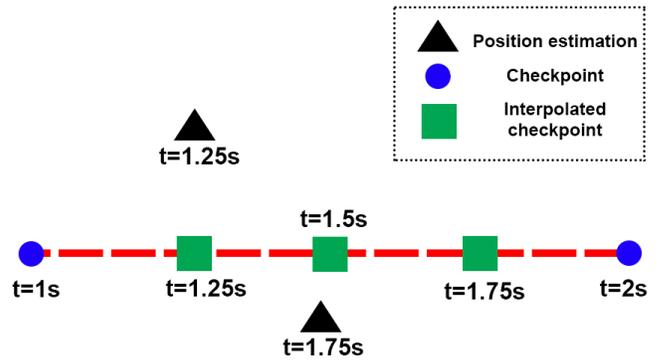


Figure 4. Linear interpolation example using fixed time interval in a two-dimensional map

C. Data evaluation

Finally, regardless of the interpolation method used, some parameters must be chosen for the evaluation of a given provider. Keeping in mind the evaluation of the end user experience we have chosen the following:

Euclidean distance error: The most popular way to evaluate a positioning system is stating the average error committed between the estimated position and the real one. Using the proposed methodology, the errors are calculated individually for every interpolated checkpoint. Therefore many common statistical parameters can be calculated with this procedure, such as:

- Mean
- Median
- 75th percentile
- Standard deviation
- Cumulative Distribution Function (CDF)

Floor accuracy: It is given by the ration between the number of correct detections of the floor in the position estimates divided by the total number of estimates. Similarly to the room accuracy, a position provider can estimate the altitude or the floor of the position evaluation, analysing the accuracy of this estimation can allow the end user to trust the veracity of this information or not.

Room accuracy: It is given by the ration between the number of correct detections of the room in the position estimates divided by the total number of estimates. Some indoor positioning providers are able to estimate the room the end user is inside of. If this parameter is added when defining the ground truth path, the accuracy of the provider when estimating the room can be calculated.

Time to first fix: It is also important to indicate the time it takes for a position provider to give the first position estimate since its initialisation.

Accuracy estimation evaluation: The accuracy estimation (AE) is another key parameter for perceived utility of the position estimates by the end user. This value can be described as the radius of $p\%$ confidence of the position, where p should be described by each provider. As an example, this value is claimed to be 68% in Android position providers, while in many GPS cases is common to utilize 95%. In other words, if a circle centred at the position estimate's latitude and longitude is drawn, and with a radius equal to the accuracy estimation, then there is a $p\%$ probability that the true position is inside the circle. The accuracy estimation provides to the end user a degree of certainty over the the position estimate received.

In the method presented in this study we attempt to have a better characterisation of the accuracy estimation by evaluating its precision during the recorded path. Using the proposed methodology, we can empirically measure p by calculating the percentage of ground truth positions inside the radius indicated by the accuracy estimation.

Travelled distance ratio: One of the most unpleasant effects for an end user when utilising a positioning system is visualising constant abrupt changes in the estimated position, as it does not transmit a feeling of continuity when walking. In order to evaluate this effect we introduce the concept of travelled distance ratio (TDR), which is defined as the ratio of the total distance from the estimated path, divided by the distance defined by the predefined path.

Ideally, this ratio should be one if the estimates were exact, as both distances would be the same. However, when abrupt changes appear, they generally deviate from the line followed by the user or go back and forward from the real position. This makes the total distance travelled by the estimated route longer than the ground truth one and, therefore, the TDR is higher than one. We consider the positioning system provides a better user experience when this ratio is closer to one. Figure 5 shows an example of a real path, in the upper part in red, and an estimated path, lower part in blue. In this case, the estimated path is twice as long as the real, therefore the TDR would be 2.

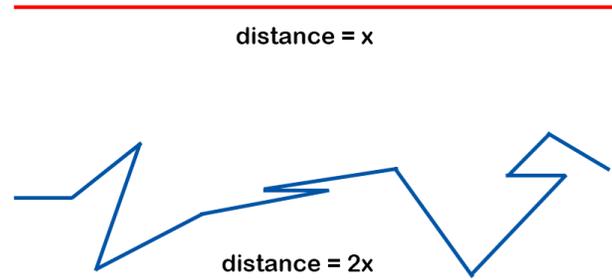


Figure 5. Example of a real path (up) and a estimated path (down)

IV. USE CASE EXAMPLE

In this section, we show an example of the full process using the proposed methodology in a real life scenario, where we desired to evaluate our positioning system in an already deployed infrastructure of Bluetooth low energy beacons, in an underground parking area. The logic of the position provider is described in the work presented by Anagnostopoulos et al. [15].

As described, the first step is to geolocalise the map of the area and create a predefined path for the tester to follow. The path is created trying to imitate how the positioning system could be used, or trying to explore a certain behaviour of it in a specific situation. We used as landmarks the lines painted on the ground for the parking spaces. The result of this procedure is shown in Figure 6.

The following step is to walk over the checkpoints, trying to maintain a steady pace and walk along the predefined path. At the same time, the position estimates given by our Bluetooth position provider are being recorded. The estimated path is shown in Figure 7.

In order to better exemplify the meaning of the statistical parameters, we also recorded, at the same time, an additional version of the same position provider. This provider performs a filter over the position estimates. In this way it smooths the variations of the position estimates in order to offer a better experience to the end user. The result of this recording is shown in Figure 8.

Once the recording step is completed, we proceed to the evaluation. For this purpose, we need to obtain the pairs of interpolated ground truth positions and position estimates for specific timestamps. In this case, we have used the fixed time interval method with a value of 250 ms.

Finally, we analyse the data recorded to extract some of the statistical parameters described in Section III. These data are shown in Table I.

It can be observed how the travelled distance ratio, has a value much closer to 1 in the case of the position provider that uses filtering techniques, and avoids abrupt changes in the position estimates. It is hard to predict the end user's

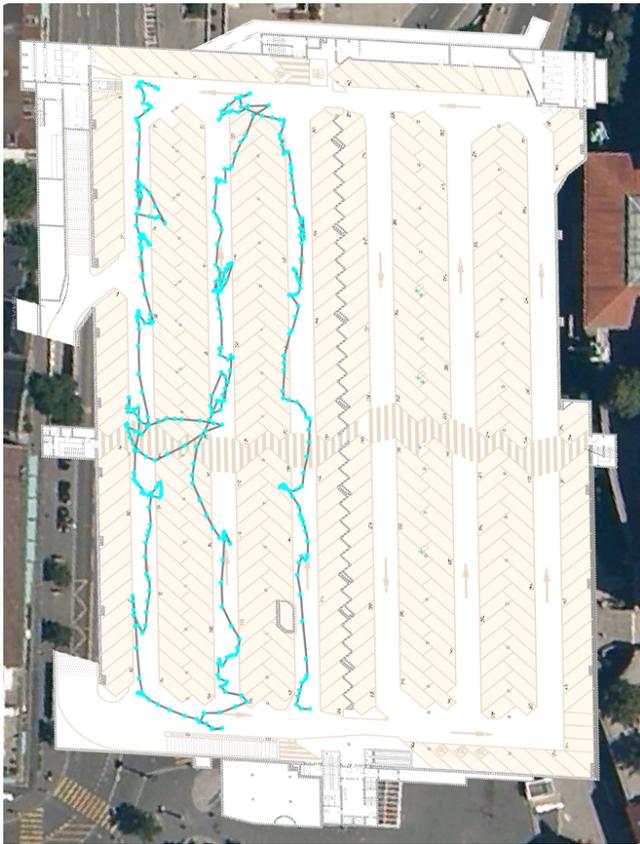


Figure 8. Estimated path by the position provider with filtering

dividing the estimated path by the real one. The closer to 1 this value is, the better we predict that the perceived utility of the system will be for the end user.

One of the main goals achieved in this work was to propose a simple and inexpensive procedure, hoping to contribute to the positioning community. Additionally, we offer a clear view of the evaluation methods used in our lab, in order to be used as reference by our future works. By using this reference, we can improve the value and credibility of the scientific process when presenting positioning research studies.

VI. FUTURE WORK

An important problem with location experiments is the reproducibility of such experiments. It is safe to assume that many positioning research groups have spent a significant amount of time walking over the same areas testing their positioning algorithms, as it has also been our case. The methodology presented in this study has a very straightforward extension, which is currently being developed in our research group. It simplifies the reproducibility problem and allows the researchers to optimize their positioning algorithms.

The idea proposed is to take advantage of the ground truth data gathering and, when recording the position estimates, record also all kinds of raw data available for the device used while walking the path. For example, if using a smartphone, it is possible to record all the Bluetooth signals received, as

well as Wi-Fi, accelerometer data, pressure data, light sensors data etc. By doing so, different algorithms or even different configurations of the same algorithm can be tested with the exact same data set in order to have consistent comparisons. Additionally, the improvement of an algorithm over time can be tested consistently.

REFERENCES

- [1] D. Lymberopoulos, J. Liu, X. Yang, R. R. Choudhury, V. Handziski, and S. Sen, "A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned," in *The 14th ACM/IEEE Conference on Information Processing in Sensor Networks (IPSN '15)*. ACM Association for Computing Machinery, April 2015. [Online]. Available: <http://research.microsoft.com/apps/pubs/default.aspx?id=241638>
- [2] 2015 *International Conference on Indoor Positioning and Indoor Navigation, IPIN 2015, Banff, Alberta, Canada, October 13-17, 2015*. [Online]. Available: <http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7336693>
- [3] T. Pulkkinen and J. Verwijnen, "Evaluating indoor positioning errors," in *Information and Communication Technology Convergence (ICTC), 2015 International Conference on*, Oct 2015, pp. 167–169.
- [4] T. Schwartz, C. Stahl, A. Mahr, C. Muller, and B. Saliba, "Using natural footstep-accurate traces for indoor positioning evaluation," in *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, Nov 2012, pp. 1–8.
- [5] D. Becker, F. Thiele, O. Sawade, and I. Radusch, "Cost-effective camera based ground truth for indoor localization," in *Advanced Intelligent Mechatronics (AIM), 2015 IEEE International Conference on*, July 2015, pp. 885–890.
- [6] S. Das, C. Gleason, S. Shen, S. Goddard, and L. Perez, "2d tracking performance evaluation using the cricket location-support system," in *Electro Information Technology, 2005 IEEE International Conference on*, May 2005, pp. 6 pp.–6.
- [7] S. Schmitt, H. Will, B. Aschenbrenner, T. Hillebrandt, and M. Kyas, "A reference system for indoor localization testbeds," in *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, Nov 2012, pp. 1–8.
- [8] T. Van Haute, E. De Poorter, F. Lemic, V. Handziski, N. Wirstrom, T. Voigt, A. Wolisz, and I. Moerman, "Platform for benchmarking of rf-based indoor localization solutions," *Communications Magazine, IEEE*, vol. 53, no. 9, pp. 126–133, September 2015.
- [9] F. Lemic, J. Busch, M. Chwalisz, V. Handziski, and A. Wolisz, "Infrastructure for benchmarking rf-based indoor localization under controlled interference," in *Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2014*, Nov 2014, pp. 26–35.
- [10] S. Adler, S. Schmitt, H. Will, T. Hillebrandt, and M. Kyas, "Virtual testbed for indoor localization," in *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*, Oct 2013, pp. 1–8.
- [11] G. Anagnostopoulos and M. Deriaz, "Automatic switching between indoor and outdoor position providers," in *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*, Oct 2015, pp. 1–6.
- [12] C. Martinez de la Osa, G. Anagnostopoulos, and M. Deriaz, "Smart position selection in mobile localisation," in *The Fourth International Conference on Communications, Computation, Networks and Technologies, INNOV 2015*, Oct 2015, pp. 44–50.
- [13] S. Adler, S. Schmitt, K. Wolter, and M. Kyas, "A survey of experimental evaluation in indoor localization research," in *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*, Oct 2015, pp. 1–10.
- [14] A. Bekkelien and M. Deriaz, "Harmonization of position providers," in *2012 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2012, Sydney, Australia, November 13-15, 2012*, 2012, pp. 1–7. [Online]. Available: <http://dx.doi.org/10.1109/IPIN.2012.6418855>
- [15] G. G. Anagnostopoulos and M. Deriaz, "Accuracy enhancements in indoor localization with the weighted average technique," in *SENSORCOMM 2014*, 2014, pp. 112 – 116.